Making Sense of Microposts at Scientific Conferences

Peter Kraker
Know-Center
Inffeldgasse 21a
8010 Graz, Austria
pkraker@know-center.at

Fleur Jeanquartier
Graz University of Technology
Inffeldgasse 21a
8010 Graz, Austria
fleur.jeanquartier@student.tugraz.at

ABSTRACT
Twitter is being widely used at scientific conferences. Following the microblogging stream, however, adds to the cognitive load of a conference participant. Therefore, there is a need for means of extracting the most important topics from a Twitter stream. This demo paper presents an adaptable system for detecting trends based on Twitter, and shows how it can be used within the setting of a conference. Following the cues of visual analytics, we use visualizations to show both the temporal evolution of topics, and the relations between different topics.

Categories and Subject Descriptors
H.1.1 [**Information Systems**]: User/Machine Systems—*Human factors*; D.2.2 [**Design Tools and Techniques**]: User Interfaces

Keywords
social media, science 2.0, visual analytics

1. INTRODUCTION
Twitter is being widely used at scientific conferences. Attendees have several reasons for using the microblogging service [2]: (1) to follow the backchannel discussion, (2) to engage with other participants and to share resources and information with them, (3) to keep up-to-date with what is going on in parallel sessions, and (4) to take down notes and send updates to non-participating followers. Nevertheless, following the Twitter stream adds to the cognitive load of a conference participant. There are too many tweets to read them all, and there is no organized way of keeping up with the backlog. This reveals the need for a means of extracting the most important topics from a microblogging stream [1].

In this paper, we present an adaptable system for detecting trends based on Twitter, and show how it can be used within the setting of a conference. The system allows for analyzing twitter streams both in real-time and in retrospect.

#MSM2012 Workshop,
April 16, 2012, Lyon, France.

Following the cues of visual analytics, our system provides two explorative visualization components: a streamgraph for analyzing topics over time and a co-occurrence network for analyzing networks of terms which may for example reveal which topics are strongly correlated. The web-based user interface[1] can be accessed in a standard web browser, but it can also be easily integrated with any system that allows for widgets adhering to the W3C standard.

The system can in principle be used to follow any Twitter stream. It is especially suitable for scientific conferences though, as it enables participants not only to detect trends in the conference stream, but also to analyze the semantic network of diverse and content-rich tweets. Moreover, being time-independent, one can later go back and review the prevalent lines of discussion at the conference.

2. SYSTEM
The system consists of three main components: (1) a Twitter crawler, (2) dataservices, and (3) visualizations. The Twitter crawler queries the Twitter API either for a list of hashtags, a list of users, or both. The tweets are logged, cleaned, and informative tokens (such as nouns, users, and hashtags) are extracted using TreeTagger[2]. With the help of cURL[3], short urls contained in the tweets are resolved, and website keywords are being extracted. Afterwards, we store the tweets, their metadata, and their associated informative tokens in a Solr[4] index. Therefore, we can both investigate real-time tweets and earlier tweets. The dataservices are REST-ful webservices, which query the Solr index, focusing either on the temporal evolution of topics or the relations between different topics. In both cases, we use Solr's facetted search capabilites to calculate the occurrences and co-occurrences of terms. The dataservices supply information to the third component: visualizations.

At the moment there are two visualizations: one is a weighted graph, a co-occurrence network for analyzing semantic networks of terms based on the JavaScript InfoVis Toolkit (JIT)[5]. The second visualization is a streamgraph based on the Grafico javascript charting library[6] for analyzing topics over time. The visualizations are embedded in an HTML-based fron-

[1] http://stellar.know-center.tugraz.at/vis/
[2] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
[3] http://http://curl.haxx.se/
[4] http://lucene.apache.org/solr/
[5] http://thejit.org
[6] http://grafico.kilianvalkhof.com/

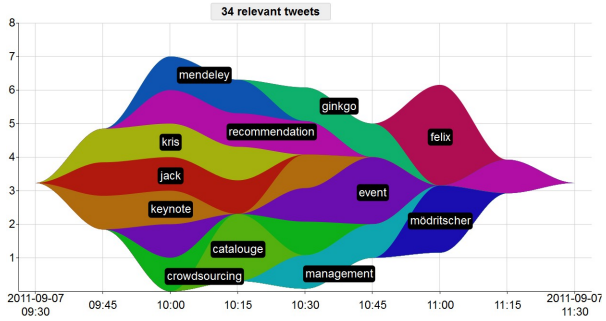

Figure 1: Streamgraph visualization from the RDSRP'11 morning session

tend that offers the possibility to type in one or several search terms, and provides a date selection to narrow the search to a specific date range. Furthermore, the UI allows for specifying a facet such as "hashtags that occurred with the term above". To allow for further inspection of the Twitter stream, the respective tweets are displayed next to the visualizations. They can be filtered for the co-occurring terms for further inspection.

3. APPLICATION TO CONFERENCES

The system can be easily adapted to individual conferences. The only prerequisite is that the hashtags relevant to the conference (e.g. #www2012 and #msm2012) are being indexed by the crawler. Then, the visualizations can be used to analyze the main topics of a conference. Figure 2 shows the weighted hashtag graph visualization of the 2nd STELLAR Alpine Rendez-vous. In the center, one can see the official hashtag for the event #arv11. The hashtags which are directly related to the event hashtag, are hashtags of individual workshops, such as #datatel11 for the dataTEL workshop, and #arv3t for the workshop "Structuring online collaboration through 3 Ts: task, time & teams". Co-occurring with the individual hashtags are hashtags that describe some of the content of the workshops, such as agency and PLE for the 3T workshop. On a metalevel, one can see that the #yam hashtag was used to inform working groups at home that use the tool Yammer.

The visualizations have been used for analysis and reflection in several workshops such as the RDSRP'11 special track at i-KNOW 2011. Figure 1 shows the streamgraph of the RDSRP'11 morning session (query: "#rdsrp11" on 07/09/2011 from 9:30 to 11:30). Each colored stream represents one co-occurring noun and its development over time. On the x-axis, the time intervals are outlined, whereas on the y-axis, the relative number of occurrences is shown. The streamgraph shows that in this session, recommendation and crowdsourcing were important topics. Furthermore, the visualization reveals that two systems were discussed (Mendeley and ginkgo).

Next to analysis and reflection, the system allows for following the conference stream in real-time. Therefore, we have developed a self-updating version of the streamgraph visualization. It queries the server in preconfigurable intervals and automatically updates the streamgraph. One way of using this visualization is to show it on large screens in the conference venue to keep participants up-to-date about the most important topics.

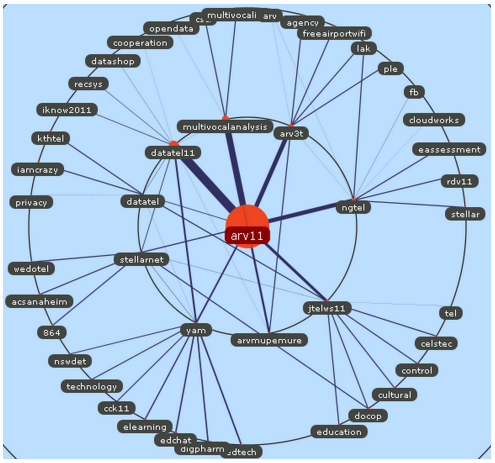

Figure 2: Weighted graph visualization of hashtags at ARV'11

4. ACKNOWLEDGMENTS

This work was carried out as part of the STELLAR Network of Excellence, which is funded by the European Commission (grant agreement no. 231913). The Know-Center is funded within the Austrian COMET program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

5. REFERENCES

[1] P. Kraker, C. Wagner, F. Jeanquartier, and S. Lindstaedt. On the Way to a Science Intelligence: Visualizing TEL Tweets for Trend Detection. In *Proceedings of the 6th European Conference on Technology Enhanced Learning*, pages 220–232, 2011.

[2] W. Reinhardt, M. Ebner, G. Beham, and C. Costa. How people are using Twitter during conferences. In V. Hornung-Prähauser and M. Luckmann, editors, *Proceeding of 5th EduMedia Conference*, pages 145–165, 2009.